Anima: Adaptive Personalized Software Keyboard

Panos Sakkos 2
panoss@microsoft.com*

Dimitrios Kotsakos 1
dimkots@di.uoa.gr

Ioannis Katakis 1
katak@di.uoa.gr

Dimitrios Gunopulos 1
dg@di.uoa.gr

[1] Dept. of Informatics and Telecommunications
University of Athens, Ilissia GR15784, Greece.

[2] Microsoft Development Center Norway
Torggata 2-4-6, PB 043, Oslo 0104, Norway.

ABSTRACT

We present a Software Keyboard for smart touchscreen devices that learns its owner's unique dictionary in order to produce personalized typing predictions. The learning process is accelerated by analysing user's past typed communication. Moreover, personal temporal user behaviour is captured and exploited in the prediction engine. Computational and storage issues are addressed by dynamically forgetting words that the user no longer types. A prototype implementation is available at Google Play Store.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining; C.5.3 [**Computer System Implementation**]: Personal Computers

General Terms

Algorithms

Keywords

software keyboards, text predictions

1. INTRODUCTION

Over the last few years smartphones and tablets, have achieved mass adoption. Currently, smart devices possess impressive computational and sensing capabilities. However, little improvement has been achieved with respect to their (software) keyboards (SKs).

Contemporary SKs are more difficult to use, due to lack of typing feedback, large number of keys and small screen size. On the other hand, the need for written communication is increasing. Popular everyday applications include chat messaging, e-mail exchanging, micro-blogging, participation in social networks and even editing documents.

With the rise of multitouch capacitive touch-screens, various research approaches [13, 2, 7] and corporate efforts [1] tried to improve software keyboards. However, none of them managed to be successfully adopted yet. Existing commercial SKs, are based on generic dictionaries [1, 5] adapting to personal typing behavior by learning new words and phrases that the user commonly uses. Generic predefined dictionaries maintain a large amount of unnecessary information and fail to capture the user language. For example, it is common for non-English speakers to type their messages in their mother tongue using Latin characters. These dialects use words that can not be found in common dictionaries, since they differ from user to user and there is no common grammar. As a consequence, 'auto-complete' solutions have disappointed users even leading them into creating memes discussing these issues [11].

We introduce `Anima`, an SK that captures its owner's unique dictionary in order to produce personalized typing predictions. `Anima` is characterized by the following *novel* features:

1. There is no requirement for a predefined dictionary
2. The learning process is accelerated by considering the user's past written communication
3. A low memory footprint is achieved, by forgetting words that the user does not use any more, and finally
4. `Anima` learns user's daily patterns and produces time-aware predictions

2. SYSTEM OVERVIEW

Architecture. According to Human Computer Interaction (HCI) research, *word* suggestion engines ('auto-complete') decrease the user's typing speed without reducing error rates [7]. Hence, we focus on predicting the *next character* that the user will most probably type. The architecture of `Anima` decouples the User Interface (UI) from the Prediction Engine (PE). In this way, we provide the opportunity to the HCI research community to develop new UI approaches that will utilize the existing PE. Hence, we defined a simple contrac that the UI must follow:

1. Inform PE about the typed character
2. Request predictions for the next character
3. Visually exploit the predicted character set that PE returns
4. Send feedback to PE when a prediction is not accurate

The system overview of `Anima` is depicted in Figure 3.

User Interface. `Anima`'s UI, is inspired by recent HCI research [6] *shrinking* the keyboard buttons that are not included in the PE's prediction set. As a result, it visually aids the user to type the desired characters (see Figure 1). In addition, shrinking buttons result to larger empty spaces between keys preventing mistypes. In order send feedback for a bad prediction the user can swipe diagonally. Finally, the user can hide the keyboard by swiping down.

*Work done while at Unversity of Athens

(a) "" (b) *"D"* (c) *"Do"* (d) *"Dog"*

Figure 2: `Anima` UI interaction while user is typing the word *Dog*

Figure 1: `Anima` *Twitter Accelerator*

3. PREDICTION ENGINE

The Prediction Engine is the module that predicts the next character that the user will type. Its main structure is based on a *Trie* data structure [12], with weights in each node. Let's assume a stream of user key strokes at each time-step t: $\{a_1, \ldots, a_t, \ldots\}$, where $a_t \in K$ and K is the set of *all* possible keys in a keyboard. A corresponding stream of prediction-*sets* is produced by the prediction engine $\{G_1, \ldots, G_t, \ldots\}$, where $G_t = \{g_{(t,1)}, \ldots, g_{(t,n_t)}\}$, with $g \in K$ and n the size of the prediction set. Note that t in n_t implies that the size of the prediction set is varying (see next paragraphs).

Learning and Predicting. Each time the user types a character a_t, PE traverses the Trie from the root, by descending one level every time a character is typed. The weights in each node, represent how many times the user has traversed this path. The weighted Trie (WT) tracks the traversed paths and adds words that the user types for the first time. A cursor (`c`) points to the current level of the Trie. Every time the user types a word separator, c is redirected to the root of WT. In order to produce predictions for the next character G_{t+1}, given the sequence of characters that have been typed after the last word separator (word W), `Anima` computes the conditional probabilities $P(k \mid W)$ for every $k \in K$, based on the weights of the nodes at c. Then, the top-n_t characters according to P are returned. The time complexity of computing all probabilities of the next character is constant, $O(1)$. This results from the following:

- The limited number of probability calculations (i.e. size of K) and
- The fact that these computations consider information that is available within the node of WT (instant access)

Constant time complexity is crucial for such applications, since any propagated delay to the UI level, would annoy the user and hurt the usability of the keyboard.

Confidence and Diffidence. An upper bound n_t is defined, representing the maximum number of predicted characters at time t. This upper bound is dynamic and is refined automatically and continuously. It is associated with the confidence of the PE. The time that n_t is adjusted, is defined by two parameters, `conf` (confidence) and `diff` (diffidence) associated with bad prediction feedback from the UI. After `conf` consecutive successful predictions, n_t is decreased by one, resulting to a more aggressive (confident) behaviour shrinking $k - n_t$ keys. On the other hand, when PE gets `diff` continuous negative feedbacks, n_t is increased by one, in order to shrink a smaller set of characters.

Learning Accelerators. In order to accelerate the learning process, *Accelerators* can be added to the PE. *Accelerators* are messages written by the user in the past. Currently one *Accelerator* has been implemented, shown in Figure 1, which crawls the user's Twitter account and feeds his/her messages to the engine, in order to train it.

Pruning. Contemporary SKs come with a pre-installed dictionary and learn words that the user is typing and are not included in the dictionary. Google's Android SK contains a dictionary of 160,722 words for the English UK language [5] and its footprint continues to increase as the user types new words (see Figure 4b). In order to address the increasing volume of the dictionary, `Anima` prunes WT, based on a *Least Recently Used* algorithm [3]. As a result, it maintains a dictionary that is more suitable to the user, since it consists of more relevant recently typed words.

Time-awareness: `Anima` incorporates Time-Awareness (TA) in the PE, in order to incorporate user daily habits into the predictions. This is done by partitioning a day into T partitions, and by keeping a WT per partition, represented as WT_t, $t \in [0, 1, ..., T]$. During a partition t, the respective WT_t is active. Our experimental results (Section 4), showed that the invocation of the time feature improved the accuracy of `Anima`.

Multilingual Support: PE is language agnostic, which means that in order to add a new language only the UI level has to be updated with the new character layout. This is a result of the way the PE handles the characters and the predictions, since WTs are independent of languages. Finally, in the case of a user typing in multiple languages, for each supported language there has to be a different instance of the PE.

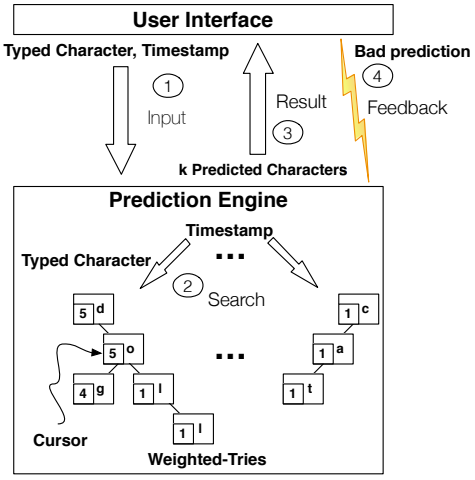


Figure 3: Anima overview

4. EVALUATION

Dataset. We utilized a dataset with messages typed through mobile devices. More specifically we collected Twitter messages and kept only the ones containing geo-location information. This information is only available when a user posts through GPS-enabled devices (smart phones, tablets, etc). We crawled Twitter using the Streaming API for the period between *February 16, 2014* and *April 6, 2014*. Tweets were collected using a bounding box around the United Kingdom, resulting in a dataset with over 27 million tweets. We selected the top 100 authors in terms of tweet count. In addition, we developed a simulator of the Anima engine (source code available here [9]), in order to simulate the users' typing behavior using Anima. Retweets and hyperlinks were ignored, because they do not contain words and characters actually typed by the users. Instead, mentions to other users were taken into account, because they can be considered as nicknames that the user uses to refer to a friend. Finally, we manually filtered-out accounts that are automated systems (e.g. weather broadcasting accounts, bots, or news broadcasting accounts). Messages were sorted in a chronological order. We simulated users typing through Anima, by using multiple-sized training sets containing 0 to 1500 tweets, with a step of 50 tweets. The test sets consist of the following 500 messages.

Results: We evaluated the predictive accuracy of Anima, by computing the hit ratio [4] over how many predicted characters were returned in each prediction. The ideal prediction engine would always predict exactly one character which would be the correct one, resulting in a precision of 1.

For the experiments, we used the two following variations of Anima:

Algorithm 1 Time-Aware typing predictions

Input:
ch: Key typed.
$feedback$: Bad Prediction Feedback.
Return:
P: List of predicted next keys along with their respective probability.

```
1: time ← clock.Now
2: if feedback then
3:    idle ← true
4:    diffidence()
5: else
6:    confidence()
7: end if
8: if idle then
9:    cursor ← TimeTries[time]
10:   cursor.Add(ch)
11:   cursor ← cursor[ch]
12:   if isWordSeparator(ch) then
13:      idle ← false
14:      TimeTries[time] ← TimeTries[time].Root
15:      return ∅
16:   end if
17: end if
18: cursor ← TimeTries[time]
19: cursor ← cursor[ch]
20: for each ch ∈ cursor do
21:    P ← P∪ popularity(ch)
22: end for
23: if isWordSeparator(ch) then
24:    TimeTries[time] ← TimeTries[time].Root
25: end if
26: return normalize(P)
```

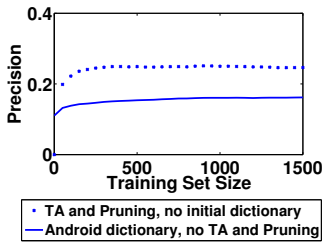
- Anima with Google's dictionary for the UK English language [5], consisting of 160,722 words, *Accelerator*-disabled, TA-disabled and *Pruning*-disabled.
- Anima with no initial dictionary, *Accelerators* disabled, TA-enabled and *Pruning*-enabled.

The dictionary approach, performed worse than Anima, which had no initial dictionary, even when using small training sets. Results are depicted in Figure 4a. The TA approaches performed better than the simple ones and as the number of the time partitions increased, the precision also increased. The results of Anima with increasing number of time partitions can be found in Table 1. Anima without *Pruning* performed worse than the version with *Pruning*, but the losses were covered from the gain of the TA.

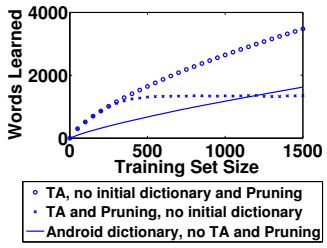
Moreover, we studied the size of the words learned when *Pruning* was disabled and we observed that the number of words was increasing with the training set size. Figure 4b depicts a comparison of the different methods with respect to the number of used words as a function of the training set size. Even for the Google's dictionary of 160,722 words, Anima continued learning new words as the training set increased, as depicted in Figure 4.

5. CONCLUSION AND FUTURE WORK

In this paper we introduce Anima, an adaptive software keyboard for touchscreen devices that learns the user's dictionary and habits, in order to produce more accurate character predictions. During the demonstration, the users will be able to test Anima with a variety of touchscreen devices



(a) Precission



(b) Words learned

Figure 4: Results

TA	Pruning	Prec	TA	Pruning	Prec
✗	✗	.2387	$T = 14$	✗	.2454
$T = 2$	✗	.2426	$T = 15$	✗	.2455
$T = 3$	✗	.2418	$T = 16$	✗	.2457
$T = 4$	✗	.2426	$T = 17$	✗	.2457
$T = 5$	✗	.2432	$T = 18$	✗	.2458
$T = 6$	✗	.2435	$T = 19$	✗	.2460
$T = 7$	✗	.2439	$T = 20$	✗	.2461
$T = 8$	✗	.2442	$T = 21$	✗	.2462
$T = 9$	✗	.2446	$T = 22$	✗	.2463
$T = 10$	✗	.2448	$T = 23$	✗	.2464
$T = 11$	✗	.2450	$T = 24$	✗	.2465
$T = 12$	✗	.2452	$T = 24$	✓	.2392
$T = 13$	✗	.2454	✗	✓	.2361

Table 1: Mean Precision for `Anima` variations, without any *Accelerator* and no initial dictionary

and settings by providing their Twitter username, so that `Anima` can accelerate learning of their personal dictionary. We plan to incorporate location features and knowledge exchanging between users that have similar behavior, e.g. users that are co-workers or are connected through a real-life relationship. Finally, we intend to use more *Acceleration* sources, like chat messages, e-mails and social networking services.

An early prototype of `Anima`, without the *Twitter Accelerator* and the TA feature, is publicly available at Google's Play Store [10], with more than 9000 installations, a rating of 4 out of 5, among 72 users and 196 Google +1s. Source code is available at [8].

6. DEMONSTRATION

Attendants will be able to use two software keyboards:

1. *Time-Aware* `Anima`, after being *Accelerated* by their Twitter account and
2. `Anima` without *Time-Aware* or *Acceleration* and with Google's Android dictionary [5] as initial dictionary

and compare them by using them.

7. REFERENCES

[1] Apple. ios 8. `http://www.apple.com/ios/ios8/quicktype/`, Apr. 2014.

[2] X. Bi, S. Azenkot, K. Partridge, and S. Zhai. Octopus: evaluating touchscreen keyboard correction and recognition algorithms via. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 543–552. ACM, 2013.

[3] W. Effelsberg and T. Haerder. Principles of database buffer management. *ACM Transactions on Database Systems (TODS)*, 9(4):560–595, 1984.

[4] N. Garay-Vitoria and J. Abascal. Text prediction systems: a survey. *Universal Access in the Information Society*, 4(3):188–203, 2006.

[5] Google. Latinime. `https://android.googlesource.com/platform/packages/inputmethods/LatinIME`, June 2014.

[6] A. Gunawardana, T. Paek, and C. Meek. Usability guided key-target resizing for soft keyboards. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 111–118. ACM, 2010.

[7] E. Rodrigues, M. Carreira, and D. Gonçalves. Improving text entry performance on tablet devices. *Interação*, 2013.

[8] P. Sakkos. Writeright source code. `https://github.com/PanosSakkos/write-right-app`, Dec. 2011.

[9] P. Sakkos. Prediction engine source code. `https://github.com/PanosSakkos/anima`, Nov. 2013.

[10] P. Sakkos. Writeright keyboard (english). `https://play.google.com/store/apps/details?id=panos.sakkos.softkeyboard.writeright`, Sept. 2013.

[11] Userbase. Damn you auto-correct! `http://www.damnyouautocorrect.com`, June 2014.

[12] D. E. Willard. New trie data structures which support very fast search operations. *Journal of Computer and System Sciences*, 28(3):379–394, 1984.

[13] Y. Yin, T. Y. Ouyang, K. Partridge, and S. Zhai. Making touchscreen keyboards adaptive to keys, hand postures, and individuals: a hierarchical spatial backoff model approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2775–2784. ACM, 2013.